

FIGURE 1

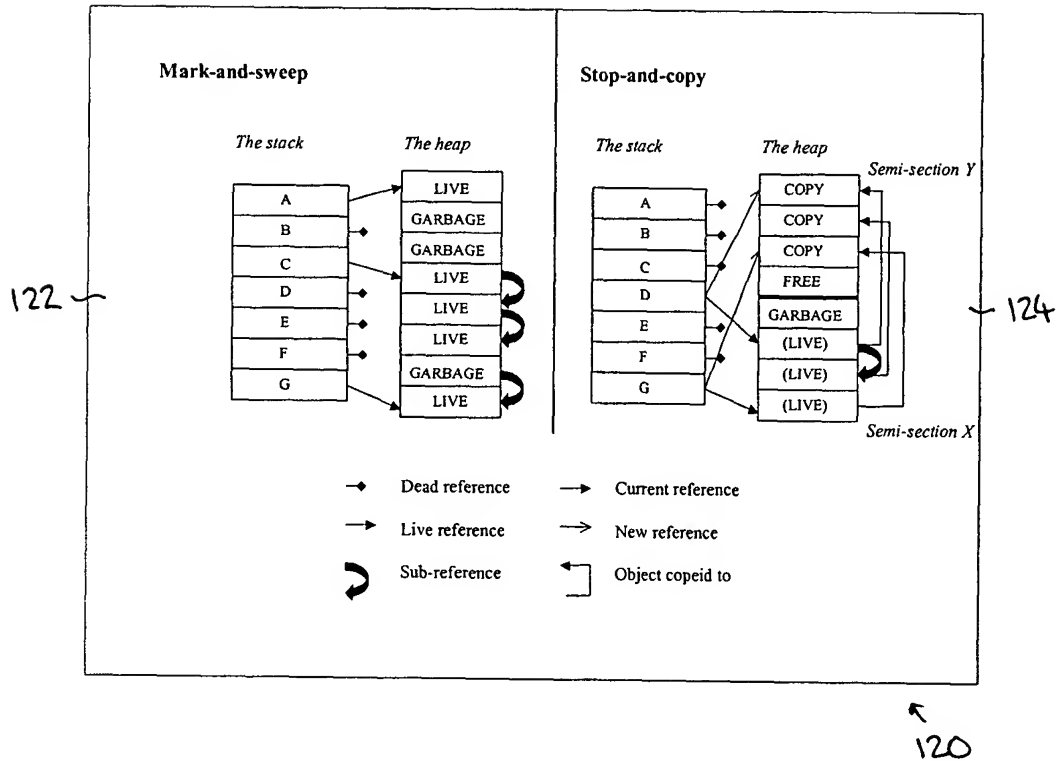
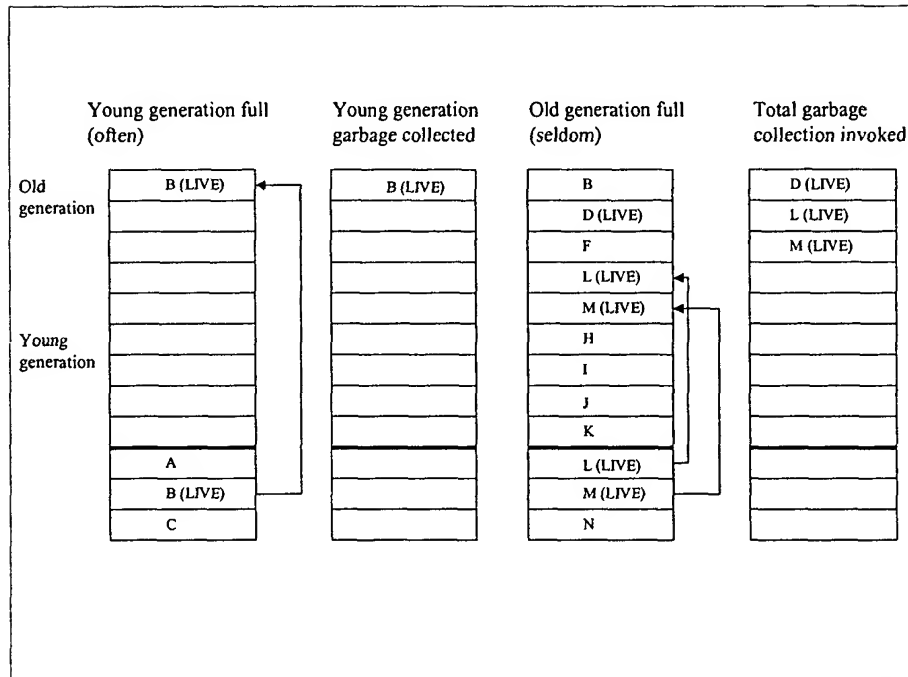


FIGURE 2



140

FIGURE 3

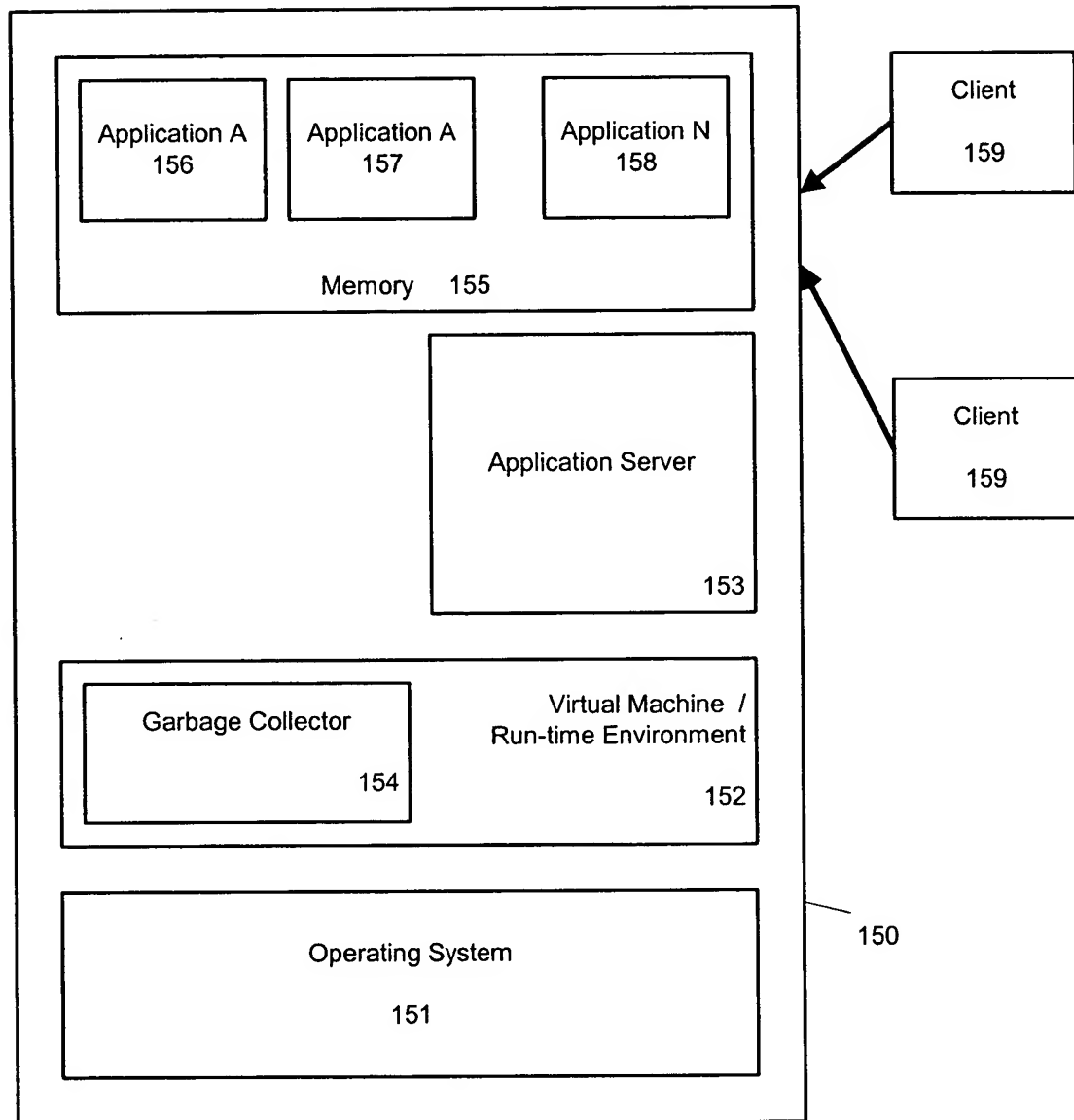


FIGURE 4

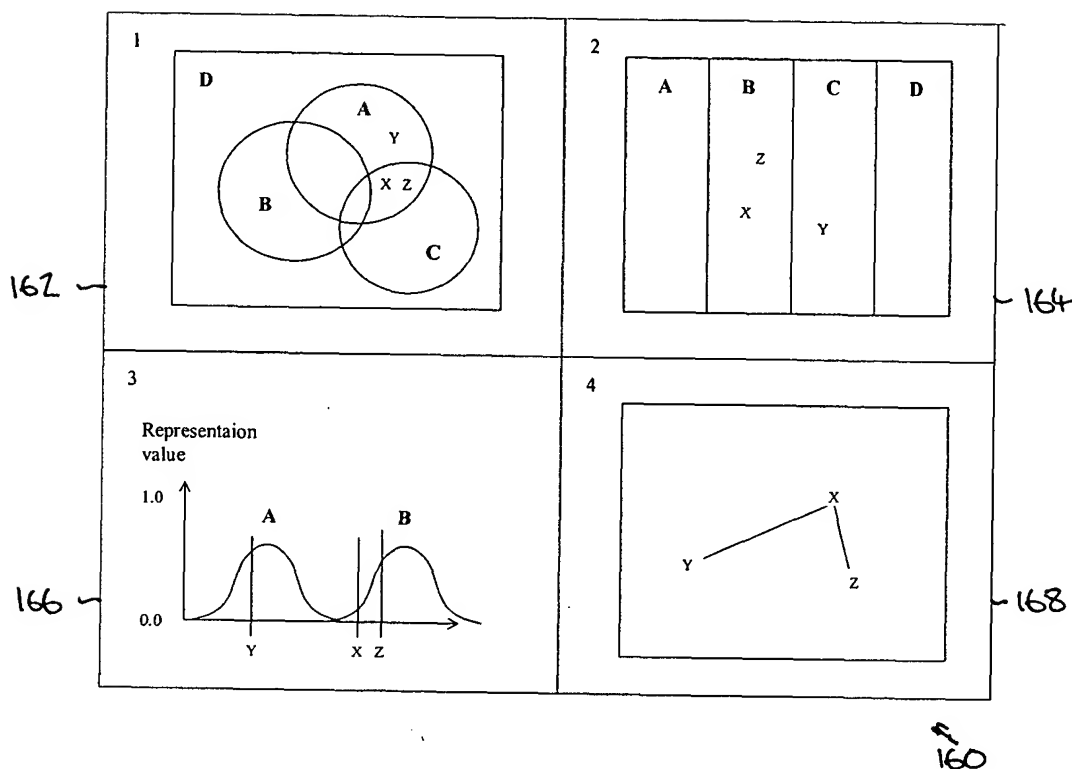


FIGURE 5

1. Environment \rightarrow State + Reward \rightarrow Decision process
2. Decision process \rightarrow Action \rightarrow Environment
3. Environment \rightarrow new State + new Reward \rightarrow Decision process

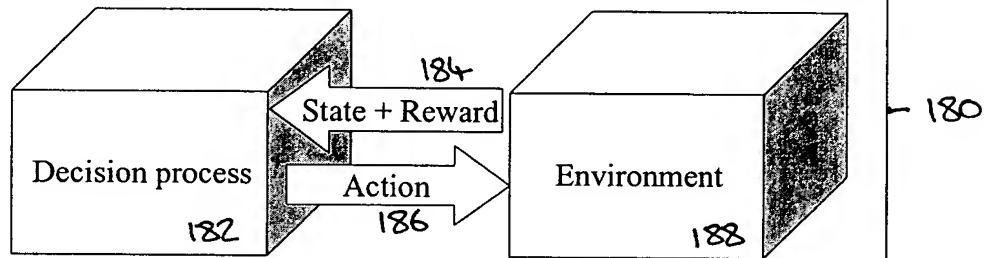


FIGURE 6

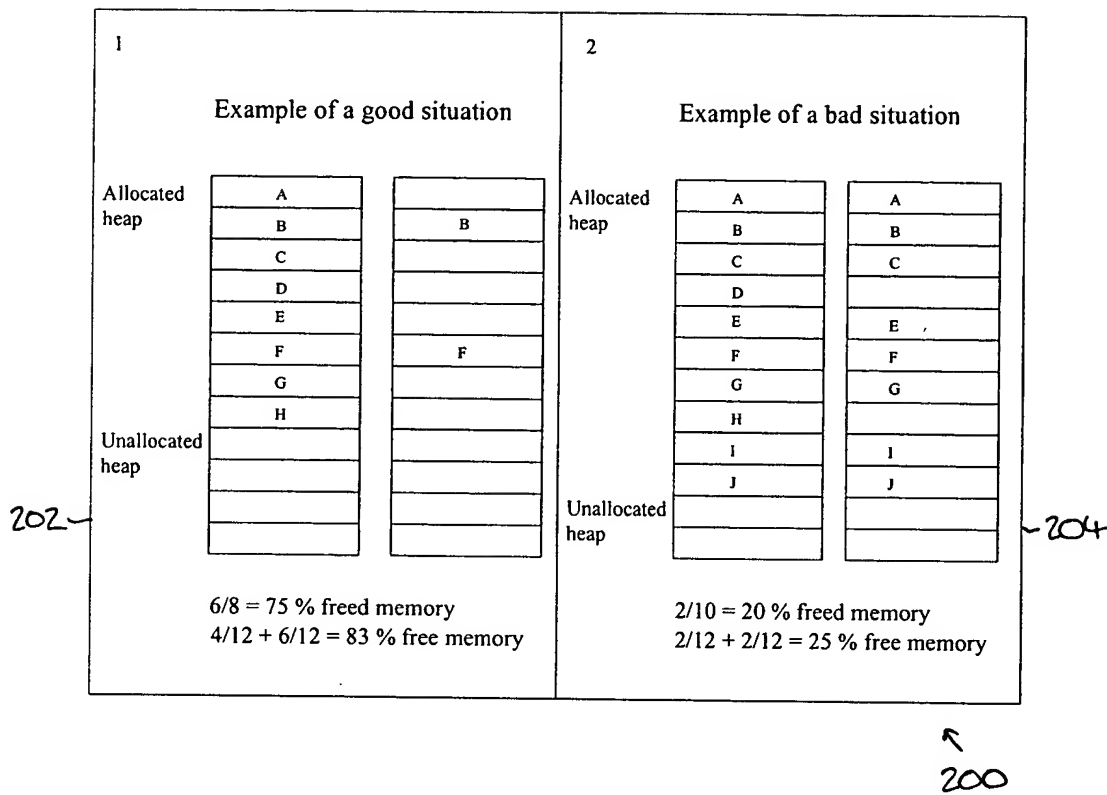


FIGURE 7

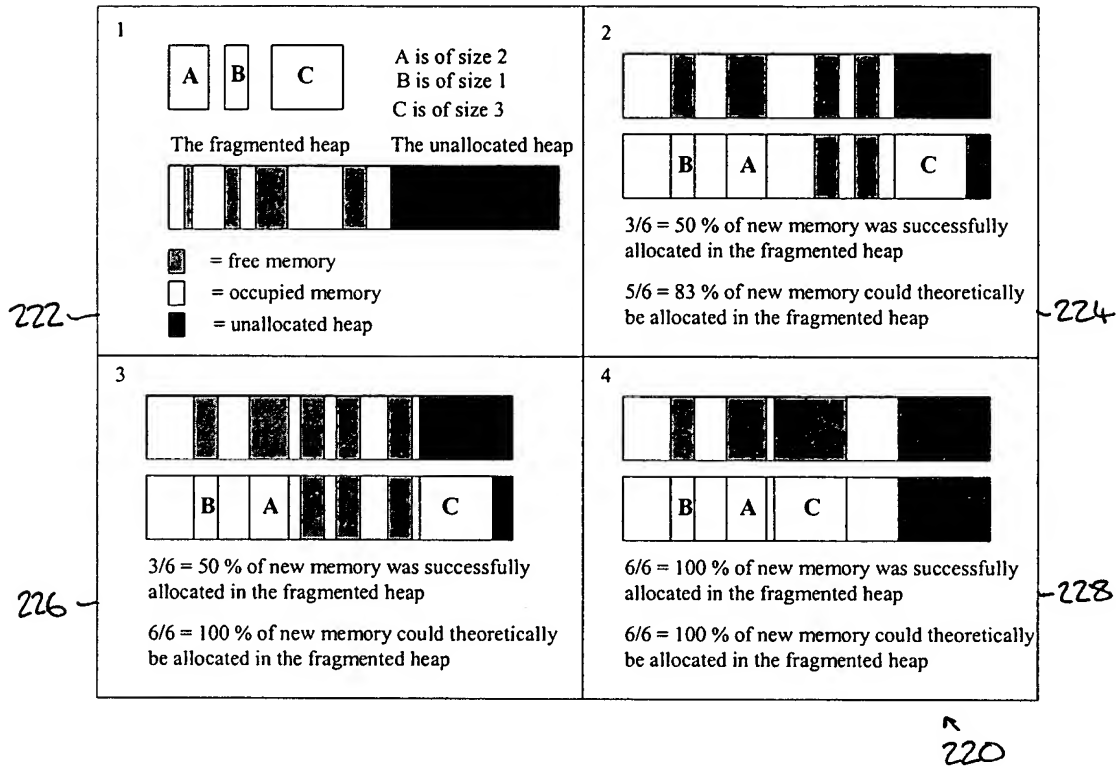


FIGURE 8

Initialize θ arbitrarily
Repeat for each episode
 $s \leftarrow$ initial state of episode
 $t \leftarrow 0$
For all $a \in A(s)$ (all possible actions a to take from state s)
 $F_a \leftarrow$ set of features present in $s = [s_1, \dots, s_n]$ and $a = [a_1, \dots, a_m]$
 $Q_a \leftarrow \sum_{i \in F_a} \theta_i$ *
 $a \leftarrow \operatorname{argmax}_a Q_a$ **
with probability $\epsilon = 1/t$: $a \leftarrow$ random action $\in A(s)$
Repeat for each step of the episode
Take action a
Observe r (the reward)
Observe s' (the next state)
For all $a' \in A(s')$
 $F_{a'} \leftarrow$ set of features present in $s' = [s'_1, \dots, s'_n]$ and $a' = [a'_1, \dots, a'_m]$
 $Q_{a'} \leftarrow \sum_{i \in F_{a'}} \theta_i$ *
 $a' \leftarrow \operatorname{argmax}_{a'} Q_{a'}$ **
with probability $\epsilon = 1/t$: $a' \leftarrow$ random action $\in A(s')$
 $\theta \leftarrow \theta + \alpha[r + \gamma Q(s', a') - Q(s, a)] \nabla_{\theta} Q(s, a, \theta)$,
where $Q(s', a') = Q_{a'}$ and $Q(s, a) = Q_a$
 $a \leftarrow a'$
 $t \leftarrow t + 1$
Until S' is the terminal state
Until eternity
* For each action-state pair there will be an estimate of the value of the pair, based on the sum of the values in θ at indexes corresponding to the features present in s, a .
** Find the highest state-action value for this state and choose the corresponding action.

240

FIGURE 9

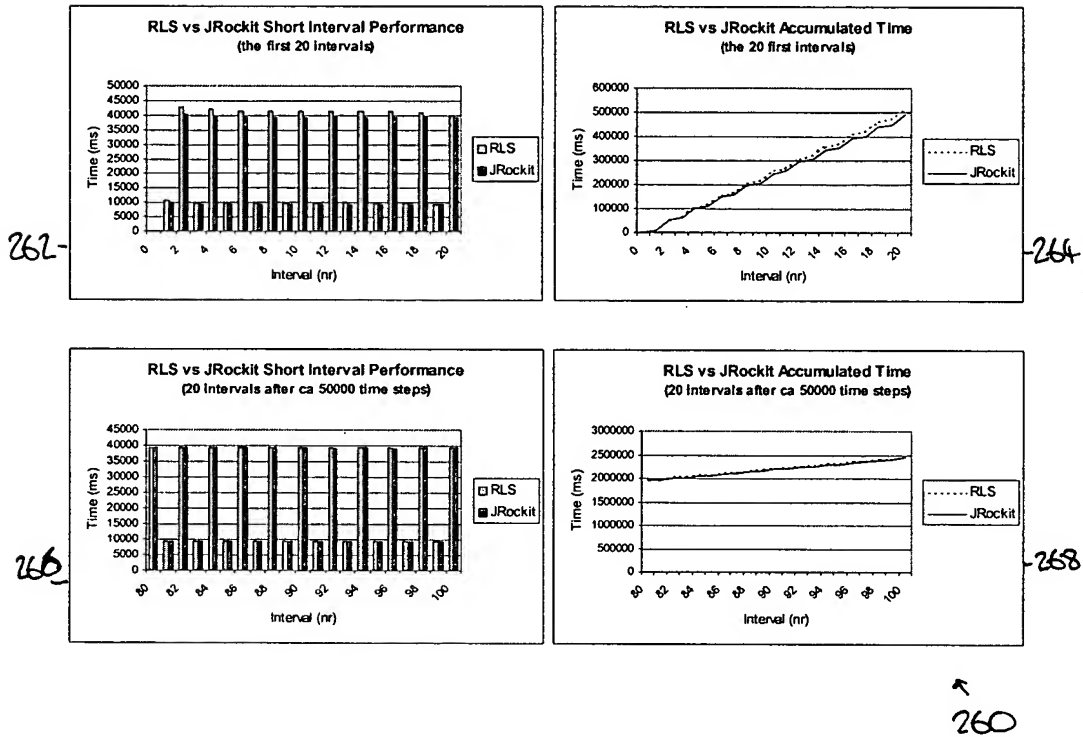


FIGURE 10

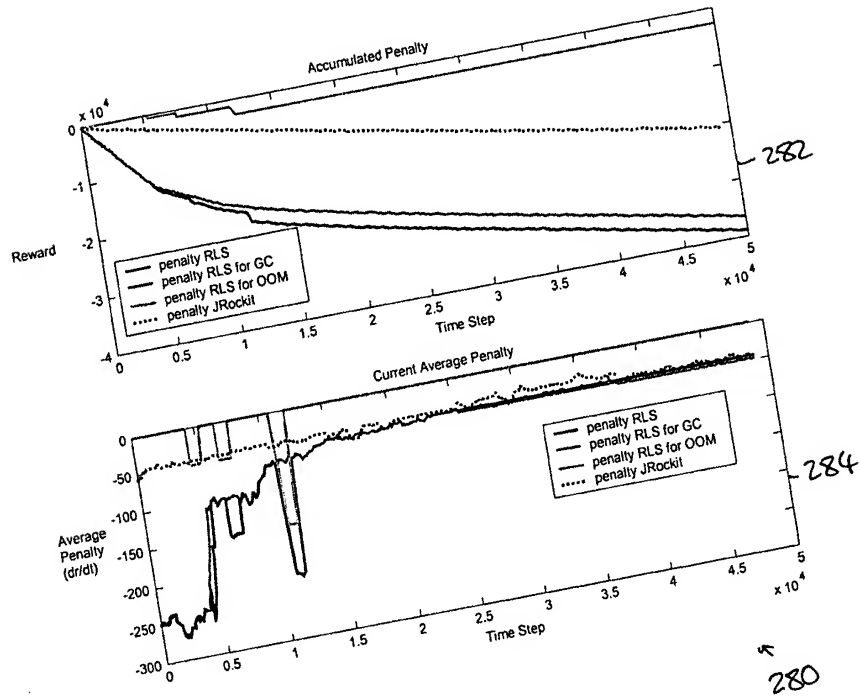


FIGURE 11

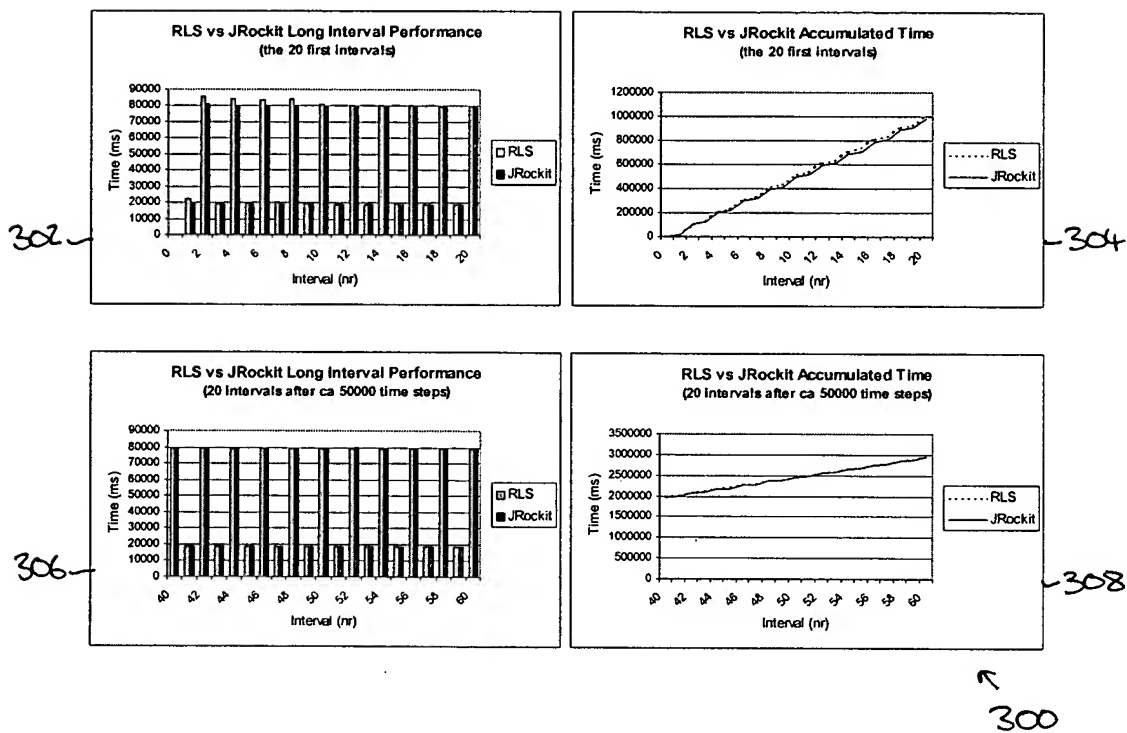


FIGURE 12

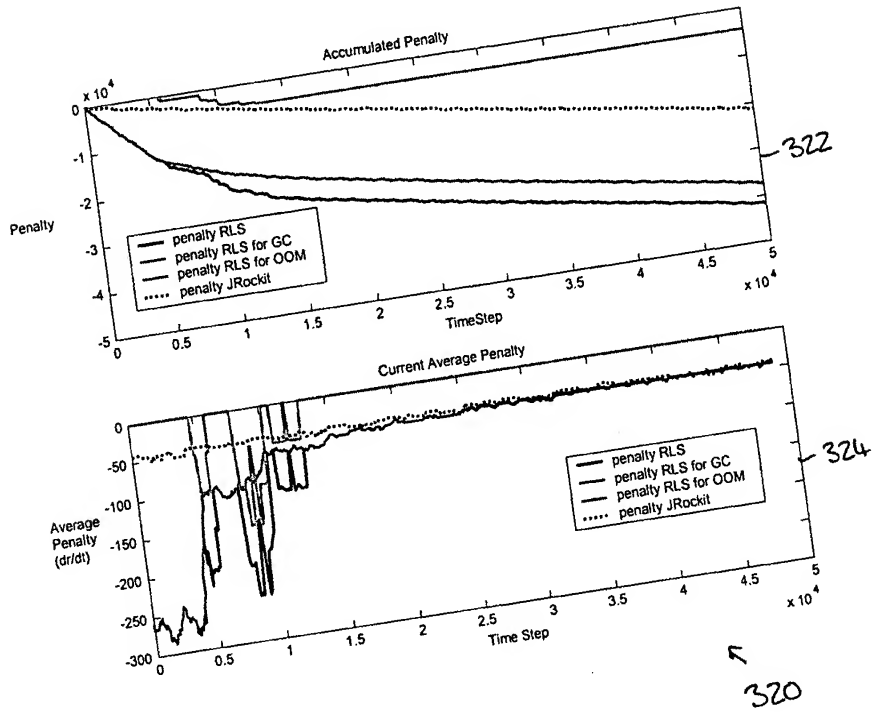


FIGURE 13

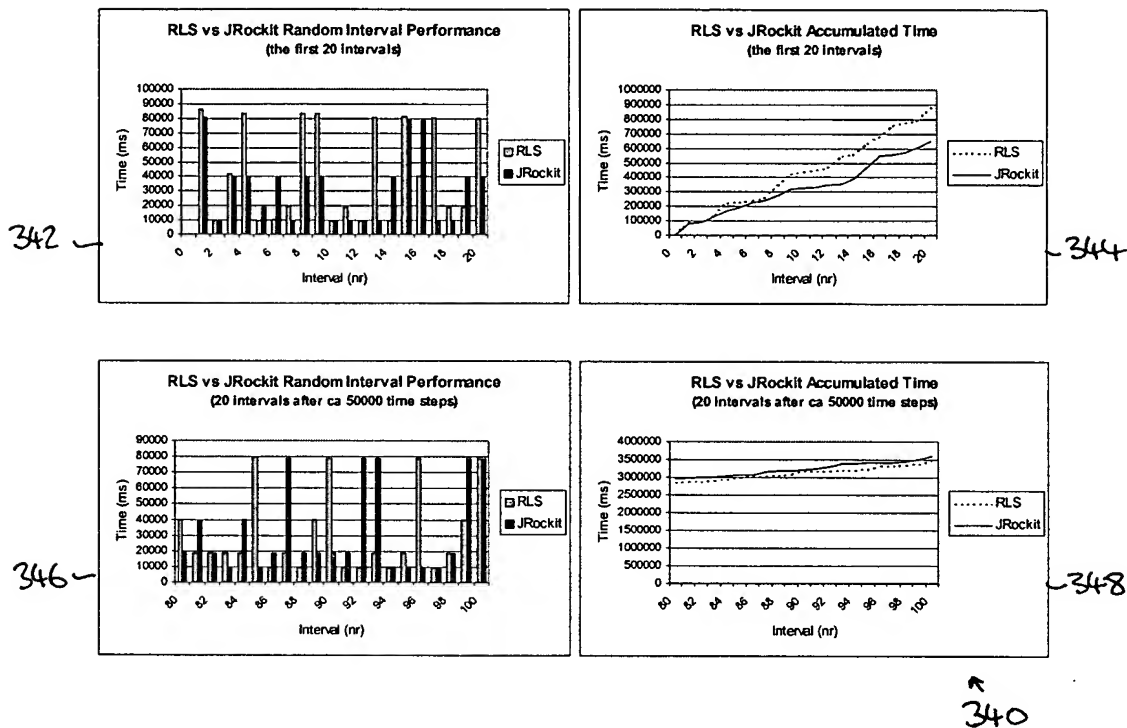


FIGURE 14

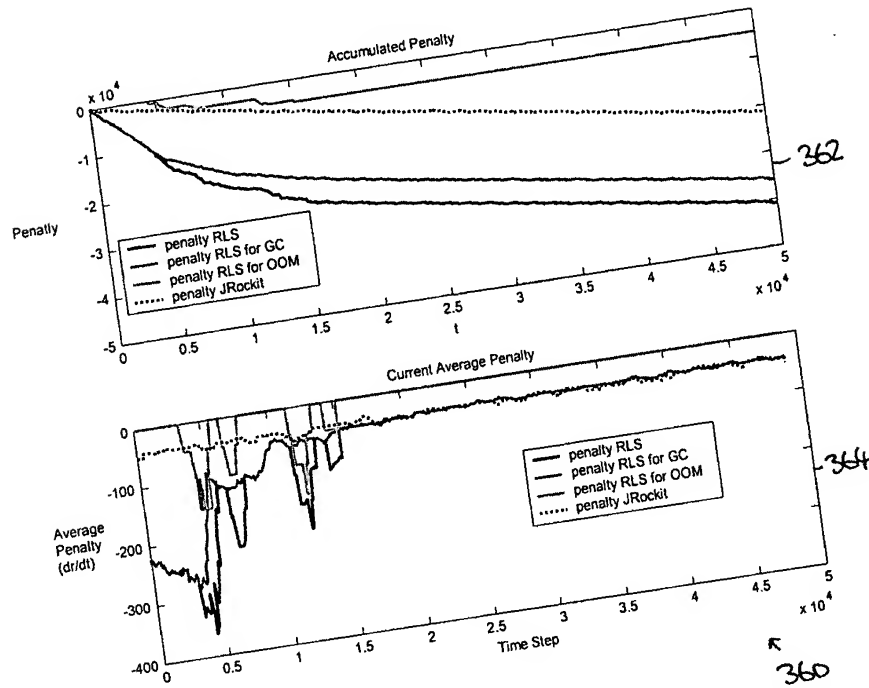


FIGURE 15

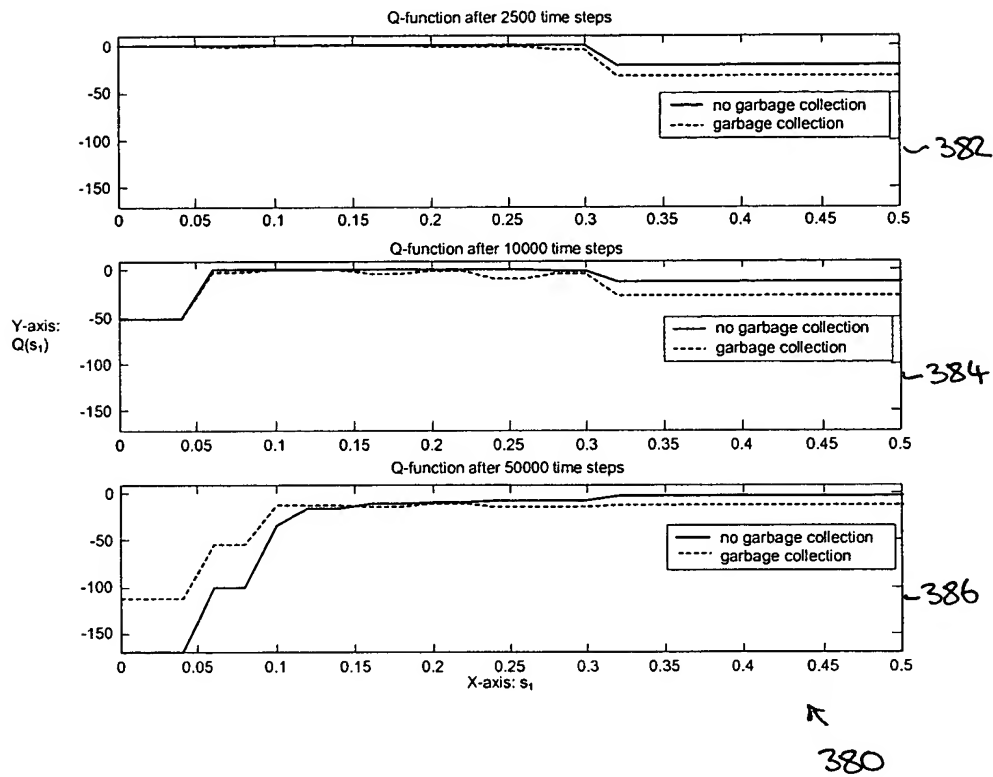


FIGURE 16

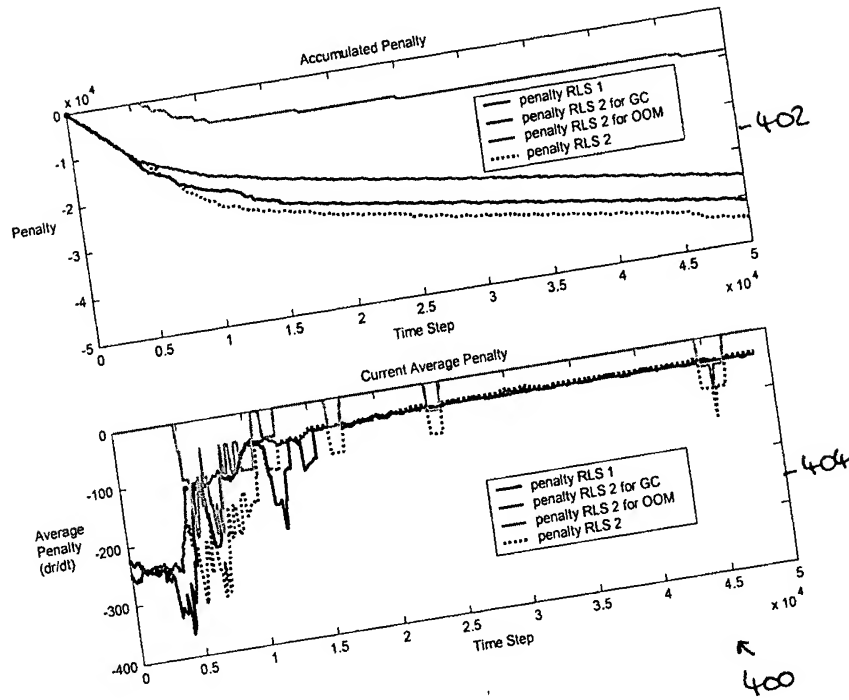
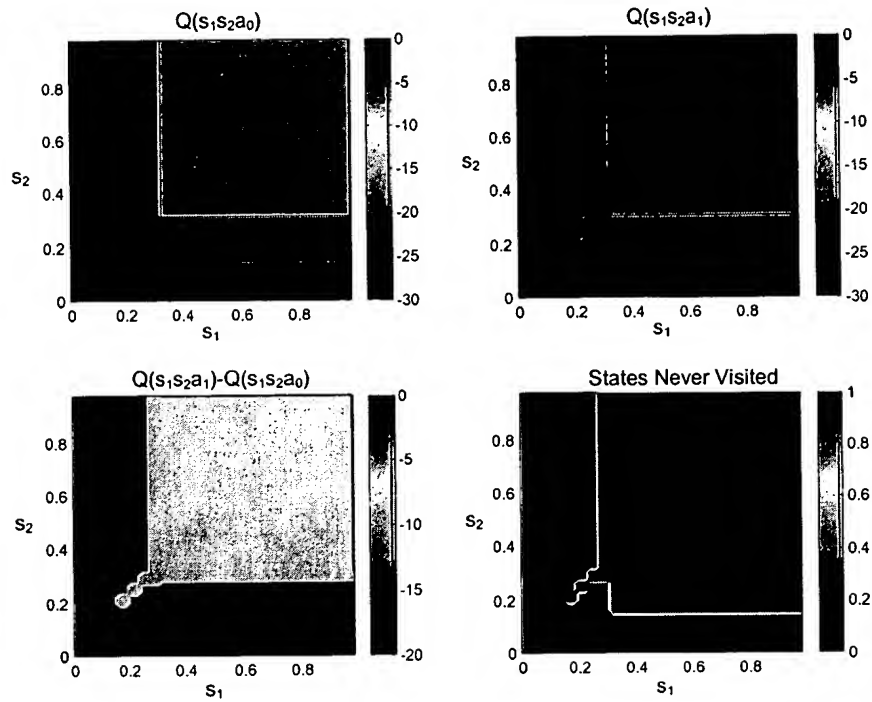
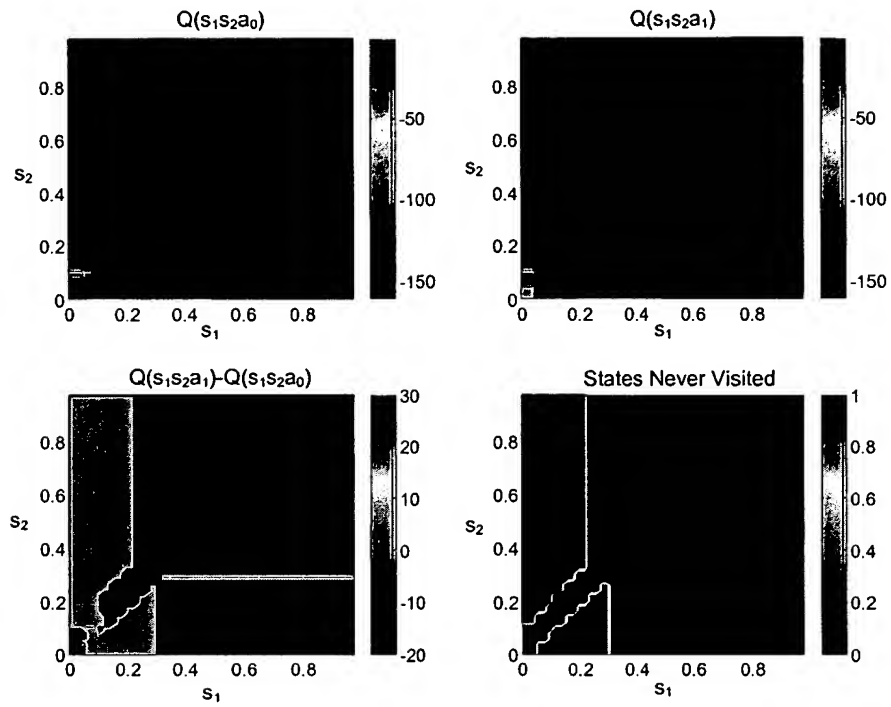


FIGURE 17



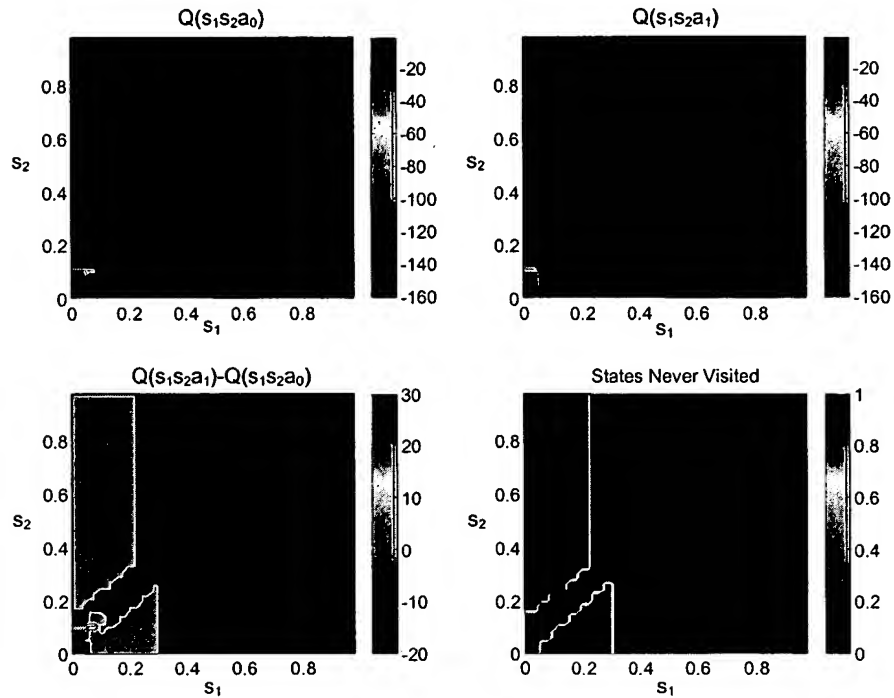
↑
420

FIGURE 18



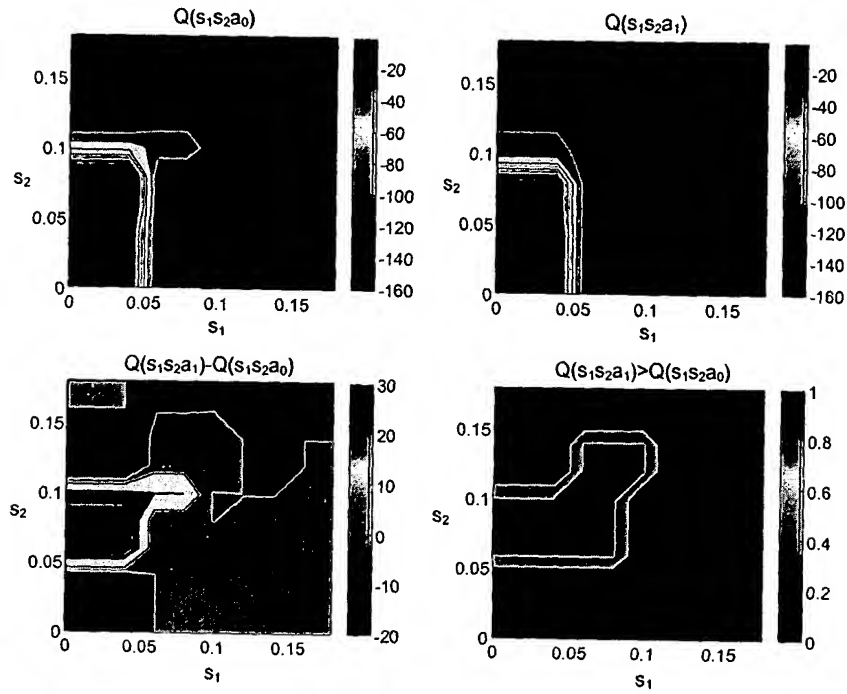
7
444

FIGURE 19



↖
460

FIGURE 20



480

FIGURE 21